Extra Credit

Examine the robustness of the *independent two sample t-test.*

What is the distribution of the p-value?

1. Simulate two independent samples:

Sample 1: $X_1, ..., X_n \sim N(\mu_1, \sigma_1^2)$, with $n = 15$, $\mu_1 = 5$ and $\sigma_1^2 = 2$
Sample 2: $Y_1, ..., Y_m \sim N(\mu_2, \sigma_2^2)$, with $m = 20$, $\mu_2 = 5$ and $\sigma_2^2 = 3$

   (a) Perform an *unequal variance independent two sample t-test*, to test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. Determine the *p-value* of the test.
   (b) Simulate repeated sampling from the two normal populations, under the null hypothesis $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. What is the distribution of the *p-value* under the null hypothesis?

2. Simulate two independent samples:

Sample 1: $X_1, ..., X_n \sim N(\mu_1, \sigma_1^2)$, with $n = 15$, $\mu_1 = 5$ and $\sigma_1^2 = 2$
Sample 2: $Y_1, ..., Y_m \sim N(\mu_2, \sigma_2^2)$, with $m = 20$, $\mu_2 = 7$ and $\sigma_2^2 = 3$

   (a) Perform an *unequal variance independent two sample t-test*, to test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. Determine the *p-value* of the test.
   (b) Simulate repeated sampling from the two normal populations, under the null hypothesis $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. What is the distribution of the *p-value* under the null hypothesis?

```
# Question 1

n = 15; mu1 = 5; sigma1 = 2;  m = 20; mu2 = 5; sigma2 = 3

x.samp = rnorm(n, mu1, sigma1)
y.samp = rnorm(m, mu2, sigma2)

xy.samp = c(x.samp, y.samp)

grp = c(rep(1,n),rep(2,m))

xy = data.frame(xy = xy.samp)
xy$grp = as.factor(grp)

attach(xy)

boxplot(xy ~ grp, data = xy)

# one-sided independent two sample t-test with unequal variance

t.test(xy.samp[grp=="2"], xy.samp[grp=="1"],
       alternative = "two.sided", mu = 0, var.equal = FALSE)

# Question 2

n = 15; mu1 = 5; sigma1 = 2;  m = 20; mu2 = 7; sigma2 = 3

x.samp = rnorm(n, mu1, sigma1)
y.samp = rnorm(m, mu2, sigma2)

xy.samp = c(x.samp, y.samp)

grp = c(rep(1,n),rep(2,m))

xy = data.frame(xy = xy.samp)
xy$grp = as.factor(grp)

attach(xy)

boxplot(xy ~ grp, data = xy)

# one-sided independent two sample t-test with unequal variance

t.test(xy.samp[grp=="2"], xy.samp[grp=="1"],
       alternative = "two.sided", mu = 0, var.equal = FALSE)
```

```
# Distribution of the p-value

B = 10000

p.val = numeric(B)

for(i in 1:B){
x.samp = rnorm(n, mu1, sigma1)
y.samp = rnorm(m, mu2, sigma2)

p.val[i] = t.test(x.samp, y.samp,
       alternative = "two.sided", mu = 0, var.equal = FALSE)$p.value

}

hist(p.val)
```

3. Simulate two independent samples:

Sample 1: $X_1, ..., X_n \sim Poisson(\mu_1)$, with $n = 15$, $\mu_1 = 5$
Sample 2: $Y_1, ..., Y_m \sim Poisson(\mu_2)$, with $m = 20$, $\mu_2 = 5$

(a) Perform an *unequal variance independent two sample t-test*, to test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. Determine the *p-value* of the test.

(b) Simulate repeated sampling from the two Poisson populations, under the null hypothesis $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. What is the distribution of the *p-value* under the null hypothesis?

(c) Perform the G.L.R. test.

4. Simulate two independent samples:

Sample 1: $X_1, ..., X_n \sim Poisson(\mu_1)$, with $n = 15$, $\mu_1 = 5$
Sample 2: $Y_1, ..., Y_m \sim Poisson(\mu_2)$, with $m = 20$, $\mu_2 = 7$

(a) Perform an *unequal variance independent two sample t-test*, to test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. Determine the *p-value* of the test.

(b) Simulate repeated sampling from the two Poisson populations, under the null hypothesis $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. What is the distribution of the *p-value* under the null hypothesis?

(c) Perform the G.L.R. test.

```
n = 15; m = 20

set.seed(1)

mux = 5
muy = 7

x = rpois(n,mux);x;length(x);mean(x);var(x)
y = rpois(n,muy);y;length(y);mean(y);var(y)

mean(y)/mean(x)

plot(y,x)
```